Softmax Regression.

$$z^{[l]} = w^{[l]} a^{[l-1]} + b^{[l]}$$

Activation Function:

$\rightarrow t = e^{(z^{[l]})}$

$\rightarrow a^{[l]} = \frac{t_i}{\sum_i t_i}$

$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$ "hard max"

"softmax"

$$z^{[l]} = \begin{bmatrix} 5 \\ 2 \\ -1 \\ 3 \end{bmatrix} \quad t = \begin{bmatrix} e^5 \\ e^2 \\ e^{-1} \\ e^3 \end{bmatrix} \quad a^{[l]} = g^{[l]}(z^{[l]}) = \begin{bmatrix} e^5/(e^5+e^2+e^{-1}+e^3) \\ \vdots \\ \vdots \\ \vdots \end{bmatrix} = \begin{bmatrix} 0.842 \\ 0.042 \\ 0.002 \\ 0.114 \end{bmatrix}$$

softmax regression generalizes logistic regression to C classes.

- ☆ if C=2, softmax reduces to logistic regression

$$a^{[l]} = \begin{bmatrix} 0.842 \\ \cancel{0.158} \end{bmatrix}$$

result from logistic regression (scalar)

Loss function:

$$y^{[i]} = \begin{bmatrix} 0 \\ \boxed{1} \\ 0 \\ 0 \end{bmatrix} \text{ - cat.} \qquad a^{[l](i)} = \hat{y}^{(i)} = \begin{bmatrix} 0.3 \\ \boxed{0.2} \\ 0.1 \\ 0.4 \end{bmatrix}$$

$$L(\hat{y}, y) = -\sum_{j=1}^{4} y_j \log \hat{y}_j \quad \Big| \quad J(W^{[1]}, b^{[1]} \ldots) = \frac{1}{m}\sum_{i=1}^{m} L(\hat{y}^{(i)}, y^{(i)})$$

$= -y_2 \log \hat{y}_2 = -\log \hat{y}_2 \rightarrow$ make $\hat{y}_2$ big to reduce loss.

$$Y = [\, y^{(1)}, y^{(2)} \cdots y^{(m)} \,]$$

$$= \begin{bmatrix} 0 & 1 & 0 & \\ 1 & 0 & 0 & \\ 0 & 0 & 0 & \cdots \\ 0 & 0 & 1 & \end{bmatrix}$$

$(4, m)$

$$\hat{Y} = [\, \hat{y}^{(1)}, \hat{y}^{(2)}, \cdots, \hat{y}^{(m)} \,]$$

$$= \begin{bmatrix} 0.2 & \\ 0.3 & \cdots \\ 0.2 & \\ 0.3 & \end{bmatrix}$$

$(4, m)$.